融合多层次数据的问答知识图谱本体模型构建*

■ 周毅[1] 刘峥[1,2] 粟小青[3] 金体成[3]
[1]中国科学院文献情报中心 北京 100190 [2]中国科学院大学经济与管理学院图书情报与档案管理系 北京 100190
[3]华为终端有限公司 深圳 518129

摘 要:[目的/意义]针对基于问答对的智能问答准确率和解决率低、用户满意度差等问题,研究构建知识图谱本体模型,构建基于知识图谱的智能问答,解决基于问答对的智能问题所面临的问题。[方法/过程]首先,分析当前智能问答面临的问题及原因,提出构建知识图谱支撑智能问答的方案。其次,在已有本体模型构建方法的基础上,提出一种融合多层次数据的多轮循环方法,该方法分别以业务数据、用户数据和业务系统动态数据等多层次数据为数据来源,核心步骤为搭建基本框架、完善知识结构、对齐知识结构三轮循环。最后,以退换货领域为例阐述本体模型构建的具体步骤,从无到有,增量叠加,构建知识图谱本体模型。[结果/结论]将以退换货本体模型为模式层的知识图谱部署在智能问答系统中进行试验,试验结果显示退换货知识图谱上线后智能问答的准确率提升50%,解决率提升300%。其中准确率是指回答正确的问题数量与回答的全部问题数量的比例,解决率是指答案精准解决了用户问题的数量与回答的全部问题数量的比例。本文提出的本体模型构建方法从零散的领域知识中梳理出完整的、细粒度的领域知识结构,支持智能问答为用户提供精准的答案,能够有效解决基于问答对的智能问答困境。

关键词:知识图谱 本体模型 精准问答 多层次数据
分类号:G250
DOI:10.13266/j.issn.0252-3116.2022.05.013

1 引言

近年来,不少企业逐渐在客服领域应用智能问答,以提高客服效率、降低企业成本。智能问答可以分为任务完成型、知识获取型和聊天型3类,其中知识获取型为本文探讨重点。知识获取型智能问答从数据组织形式角度划分有3种:①基于问答对的智能问答,事先拟定一定数量的问答对(QA对),采用关键词匹配的方式找到与用户问题匹配的问答对,选取匹配度最高的作为答案返回;②基于知识图谱的智能问答,它以知识图谱中的结构化知识为内容,支持基于语义理解的精准问答;③基于阅读理解的问答,该类问答给定文档库,通过机器阅读理解用户的问题,从文档中找到能够回答用户问题的片段,没有预先的知识抽取工作。

当前的企业智能客服多为基于问答对的智能问答。基于问答对的智能问答解决问题的能力依赖问答对的丰富程度以及人工配置相似问的情况,在实际使用中问题解决率低、用户满意度差,需要转型升级、寻找突破。基于阅读理解的智能问答仍处于研究探索阶段。基于知识图谱的智能问答支持基于多粒度层次上语义的问答匹配,并可以支持实体间上下文会话识别与推理。因此,基于知识图谱的智能问答是解决基于问答对的智能问答所处困境、实现智能客服转型升级的有效路径。

知识图谱分为模式层和实例层。开放领域的知识图谱可以只有实例层没有模式层,但垂直领域的领域知识图谱构建,对知识的质量和准确度要求高,需要领域知识图谱具有完备的模式层,以抽象出领域中的重要概念和概念之间的关系,供领域间复用与融合。因此本体模型是领域知识图谱构建的关键。本文重点探讨基于多层次领域数据的知识结构化,探索构建结构合理、覆盖全面的知识图谱本体模型。

* 本文系国家重点研发计划项目"先进制造业分布式科技服务技术集成研发与示范"(项目编号:2019YFB1405100)研究成果之一。
作者简介:周毅,馆员,硕士;刘峥,研究馆员,博士,通信作者,E-mail:liuz@mail.las.ac.cn;粟小青,知识管理专家,硕士;金体成,知识管理工程师。
收稿日期:2021-07-06 修回日期:2021-1105 本文起止页码:125-132 本文责任编辑:王传清

2 基于问答对的智能问答困境

基于问答对的智能问答面临的问题主要体现在知识的横向覆盖率、纵向细粒度及知识质量等方面。

2.1 知识覆盖率难以提升

大量的问答对松散孤立,没有关联,存在重复冗余和知识缺失的问题。一方面,运营人员需要源源不断地添加新的问答对和相似问,知识管理任务重,无法将所有问题都添加为问答对;另一方面,因缺少完整规范的领域知识体系结构,即使不断添加单个知识点,也无法确定满足用户需求的知识边界。

2.2 知识粒度难以细化

实际应用中,用户问题的概括程度不一,问答对无法准确把握粒度。如用户可能提问"退货政策是什么?",也可能提问"新买的某款产品已经过了七天能不能退?"。如果将问答对粒度设置过细,人工工作量大,难以实现。如果将问答对粒度设置过粗,答案过于概述,又不能解决用户实际问题,如用户问附近的实体店地址,答案仅提供查询实体店地址的方法,或用户询问手机电池坏了是否保修,答案提供的却是整个保修政策。

2.3 知识质量难以确保

基于问答对的智能问答需要人工编写大量的问答对及相似问,问答对的编写无法进行细致的规范控制。理解偏差、人员更替等都可能造成问答对交叉冗余,质量参差不齐,甚至会出现答案矛盾的情况。当领域知识更新时,也难以实现相关问答对的同步更新,维护工作量大。

3 基于问答对的智能问答困境解决方案

为解决以上问题,本文基于多层次领域数据,探索构建结构合理、覆盖全面的知识图谱本体模型。目标是推动基于问答对的智能问答转向基于知识图谱的智能问答,提高知识的覆盖率和知识质量,支持回答多粒度的用户问题。

3.1 以本体模型刻画领域知识结构全景

知识图谱是一种用图模型来描述知识、建模万物之间关联关系的技术方法[1]。其对领域知识的规范描述可保持知识粒度的一致性,支持知识在多个应用场景下的同步更新、语义推理,从而提高知识质量,降低维护难度。

模式层是知识图谱的概念模型和逻辑基础,对实例层进行规范约束。在模式层,节点表示概念,边表示概念关系。实例层存储模式层中类或关系的实例数据,在实例层中事实以"实体 – 关系 – 实体"或"实体 – 属性 – 属性值"的 RDF 图或属性图存储[2]。构建本体模型作为模式层有利于支持各细分业务领域知识图谱融合与复用,同时避免冗余,形成领域结构合理,覆盖全面的知识结构。

要构建知识图谱本体模型,关键在于如何对零散的领域知识进行抽丝剥茧,构建出语义知识网络全景。最早的本体构建方法研究开始于 20 世纪 90 年代。斯坦福大学 T R. Gruber 在 1993 年发表的论文中讨论了基本的本体设计标准[3]。随后,M. Gruninger 和 M S. Fox[4]提供了一个基于能力问题(CQs)的本体构建的框架方法。M. Uschold 和 M. king 提出了包含识别本体目的、构建活动、评估、文档化等 4 个主要活动的本体构建方法[5]。此后的本体构建方法多是在此为基础发展而来,如 Methontology[6]、OntoKnowledge[7]、NeOn[8]、UPON[9]等。这些方法往往在领域本体构建项目中提出,步骤复杂,层层嵌套,使用时学习成本高,可复用性不强,不适用于本文研究的企业小规模业务领域本体模型构建。2001 年由 N F. Noy 和 D L. McGuiness 编写的本体 101 指南[10]阐述了本体开发的经典思想和步骤,被称为"七步法",对本文知识图谱本体模型的构建具有指导意义。鉴于传统方法的复杂性,研究人员开始在可重复和轻量化方向上探索新的本体构建方法。一方面为充分利用已有本体资源,提高本体构建效率,A. Gangemi 等[11-12]提出基于模式(OP)的方法构建本体,主张通过复用已设计的成熟本体模式构建本体。此方法充分利用前人研究成果,避免重复劳动,为知识的融合对齐奠定了基础。另一方面,UPON Lite 方法[9]在 UPON 方法的基础上轻量化本体构建方法步骤,使用表格工具开展本体构建,支持快速建立试验本体的原型。其轻量化的步骤,不依赖于复杂工具的优点,适用于企业业务领域本体模型构建。

国内的学者在多个领域开展了本体构建研究,多借鉴国外的本体构建方法[13],如朱妍昕等采用 NeOn 方法构建了哮喘药物治疗文献本体。有的则在国外方法的基础上加入新的内容,形成新的观点,如刘琳娜等在"七步法"基础上,提出新的本体构建方法,强调本体计划和循环修正,完善了原有方法。此外,不少学者尝试将叙词表转化为本体[14-16],其中贾君枝提出将叙词表转化为本体的基本原则。叙词表中已包含专业领域术语和相关关系,术语质量高,是转换成本体的优质资源。但此方法只适用于拥有成熟叙词表的特定领

域,而企业业务领域鲜有成熟的叙词表。

综合分析国内外已有的本体构建方法,发现已有的方法不能解决本体模型如何准确定位领域横向范围和纵向粒度的问题,轻量化、资源消耗低的方法更加适用于企业业务领域本体模型构建。本文借鉴本体构建101 指南、模式复用及 UPON Lite 思想,尝试从数据角度完善本体构建方法。

3.2 融合多层次数据优化覆盖率和细粒度

随着数字化进程的深化,领域知识形态更加多样,本体模型构建有了更加丰富的潜在数据源。本文将探讨融合多层次数据用于本体模型构建的方法,以期准确地刻画领域知识全景。

构建知识图谱本体模型的关键在于如何确定领域知识的横向边界和纵向粒度。早在 1995 年, M. Grüninger 等就建议采用能力问题确定本体的领域和范围[4]。其后,多种本体构建方法中提出以能力问题来确定本体涉及的范围,如本体将涉及的领域是什么?我们将使用本体做什么? 对于哪些类型的问题,本体中的信息应该提供答案? 谁将使用和维护本体等。能力问题可以在一定程度上帮助定义本体模型的总体领域和范围,可在构建前用于需求确定,但能力问题的数量有限且由少数专家人为设定,无法在细节上确定领域的边界和粒度。近期 D. Wiśniewski 等提出通过自然语言层面的词法句法分析发现更多的能力问题模式,支持本体构建[17],但对于需要支撑智能问答的知识图谱本体构建,仍然不能满足需求,故无法依赖能力问题来确定知识图谱本体模型的覆盖率和细粒度。

在数据来源上,以往的研究在构建企业相关的领域知识图谱时多采用百科数据、企业基本信息数据、企业新闻、上市数据、业务文档等静态知识构建知识图谱[18－19]。此类静态知识是知识图谱构建的重要来源,然而企业作为业务运营方提供的数据往往与真实用户需求存在差距,不足以支持构建覆盖全面、粒度合理的知识图谱本体模型。

在领域数据多样化的趋势下,利用多层次数据构建知识图谱本体模型成为可能。其中用户作为需求方提供的问题、搜索词等数据也是领域知识的重要组成部分。将用户数据用于构建本体模型,可以基于客观数据准确梳理出领域知识的横向边界。另一方面,用户在获取服务时通常需要与业务系统进行交互,将业务系统中的动态数据纳入知识图谱中可以实时定位用户的具体情况。因此,在构建知识图谱本体模型时需要融合来自业务运营方与用户两种角色,融合静态知识与动态知识两种形态的数据。其中企业提供的知识除知识文档外,还包括常见问题(frequently asked questions, FAQ)。常见问题列表一般由企业精心组织,具有提问频率高、质量好、通俗易懂的特点,是构建本体模型起步的优选数据。

4 融合多层次数据的知识图谱本体模型构建方法

融合多层次的本体模型构建方法在沿用 UPON Lite[20] 方法的基础上,运用多轮循环迭代的轻量级本体构建方法,采用业务运营方数据、用户数据及业务系统动态数据等多层次数据构建知识图谱本体模型。

本研究将本体模型构建分为准备阶段和构建阶段。准备阶段确定领域的范围和边界,以应用目标为指导收集领域知识。构建阶段分为多轮循环,每轮循环采用不同层次的领域知识作为输入,以增量迭代的方式完成本体模型构建。其中,第一轮循环以简短易理解的 FAQ 为起步数据构建基本框架;第二轮循环融合知识文档及用户数据完善本体模型;第三轮循环融合业务系统数据进一步完善,支持知识图谱与业务系统的动态交互。每个循环中包含同样的细分步骤,包括构建领域术语表、定义类和类的层次结构、定义属性、本体模型表示。每个步骤均可在电子表格中完成。如图 1 所示:

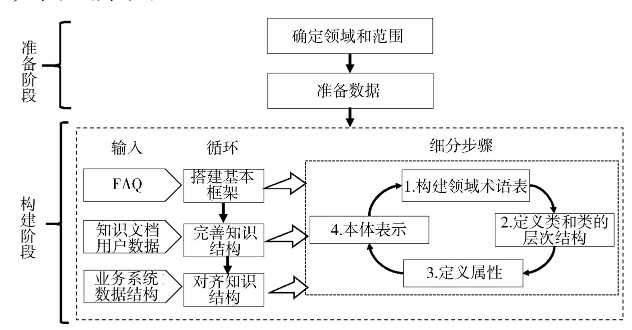


图 1 融合多层次数据的本体模型构建流程

具体流程包括:①确定领域和范围:定义知识图谱涉及的领域和范围,明确建立新知识图谱的原因、预期用途和用户类型。②准备数据:调研和搜集与本体模型构建相关的所有知识,包括可复用本体和其他知识组织资源以及可用于构建本体的领域知识。③搭建基本框架:第一轮循环以 FAQ 为起步数据搭建基本框架,包括 4 个步骤:第一步,构建领域术语表:从常见问题中识别出领域术语,构建领域术语表;第二步,定义类和类的层次结构:从领域术语中识别出独立对象及

它们之间层次关系；第三步，定义属性：描述类的内部结构；第四步，本体表示：记录完整的本体结构。④完善知识结构：融合知识文档及用户数据完善知识结构，开启第二轮循环。参照基本框架构建步骤，识别领域术语，补充类和属性，完成类和属性定义；⑤对齐知识结构：融合业务系统相关数据，进行数据结构对齐，添加类和属性，完成第三轮循环。

搭建基本框架、完善知识结构、对齐知识结构是融合多层次的领域知识构建本体模型的主要流程。在 3 个大循环下可能存在多次调整，直到构建起符合应用需求的本体模型。接下来本文将以退换货服务为例，详细阐述知识图谱本体模型的构建方法。

4.1 确定领域和范围

本体模型的构建应始终以满足应用需求为目标，没有必要包含领域的所有信息。在开始构建本体模型之前，需确定知识图谱的基本领域和范围。领域范围的界定可按照知识图谱需求说明书[21]的形式表达。

基于问答对的智能问答在退换货领域表现欠佳。用户的反馈显示，大量问题答非所问，部分问题即使命中已预设的问答对，用户仍表示不满意。为解决现有问题，本文计划构建退换货领域知识图谱，通过基于知识图谱的智能问答提高退换货领域智能问答的准确率和解决率。

4.2 准备数据

本体模型构建的原则是尽量复用已有本体或本体模式，避免重新“发明轮子”。本研究将调研和搜集领域知识图谱构建相关的所有知识，以支持知识图谱本体模型构建。一方面调研现有的本体、词表、术语表、分类表等知识组织资源；另一方面搜集退换货领域的多层次领域数据。

本文案例中，本体设计模型 OPAL（Object, Process, Actor modeling Language）[22] 支持业务本体的构建，作为知识组织资源收集备用。领域知识方面将收集退换货相关的 FAQ、知识文档、真实用户问题以及相关业务系统数据结构等作为备用。

4.3 以 FAQ 为起步数据搭建基本框架

以 FAQ 为起步数据搭建基本框架，首先要构建领域术语表，然后在领域术语表中识别作为类和属性的术语，并对类和属性进行详细定义。

4.3.1 构建领域术语表

从 FAQ 中获取领域术语，这些术语可能是本体模型要描述的对象，可能是对象的属性，也可能是属性的取值。在识别领域术语时，一是要沿着“水平”方向寻找本体模型的范围边界；二是要以应用目标为导向，沿着“垂直”方向考虑适当的细节级别或粒度。

获取领域术语后，需要添加每个术语的语义描述，构建领域术语表。添加的语义描述内容包括识别领域术语的同义词、添加文本描述、识别术语在本体模型中的类别（区分术语是类或属性）等。为术语添加语义描述的目的是对领域术语进行初步整理，以备后续步骤定义类和属性。

从退换货领域 FAQ 中获取领域术语，如“包装盒”“审核”“保修卡”“发票”等。OPAL 本体设计模型旨在通过提供有限数量的抽象概念模板支持业务本体的建设。它将类分为客体（object）、流程（process）和执行者（actor）3 种。属性类型分为填值属性（atomic property，AP）、对象属性（reference property，RP）和复杂属性（complex property，CP）3 种。退换货业务作为一种业务适用于 OPAL 模型。因此术语在本体模型中的类别以 OPAL 模型中的 3 种类和 3 种属性作区分。

4.3.2 定义类和类的层次结构

领域术语表不能表示术语间丰富的结构。本步骤将领域术语表中的独立对象进行基于专门化关系的分类，定义出类的层次结构。本文通过识别独立对象之间的 isA 关系、整体－部分（partOf）关系定义类的层次结构。

定义类和类的层次结构有自顶向下、自底向上以及两者相结合的方法。本文采用两者相结合的方法，首先定义最常见、最有把握的个别类，然后对这些类采用自顶向下和自底向上的方法进行专指化和泛化。

退换货领域应用 OPAL 的高级概念模板：客体（object）、流程（process）和执行者（actor）构建退换货领域抽象框架。在“Thing”顶类下设“Object”“Actor”“Process”3 个分支作为一级类，表示退换货业务中 3 类角色的抽象，在分支下再设下级类。

4.3.3 定义属性

属性是类的描述，定义了类和类的等级结构之后，继续从领域术语表中提炼类的属性，并连接到它们所表征的类，例如“产品”类的属性有“价格”“型号”等。属性具有各种分面特征，如属性的类型、属性的定义、属性别名、属性值类型、基数约束、单多值（属性值的数目）等。

退换货本体模型中的属性沿袭了 OPAL 模型中的属性定义，分为填值属性（atomic property，AP）、对象属性（reference property，RP）和复杂属性（complex property）3 类，并扩展了复杂属性的形式。其中，填值属性指

以数值为属性值的属性，可以直接填值，例如属性“定义”直接填字符串。对象属性指以另一个实例为属性值。OPAL 模型中关于复杂属性的定义是有内部结构并具有组成部分，例如由街道、城市、邮政编码和州组成的地址。在此基础上扩展了复杂属性的两种结构，解决复杂条件下知识的表示：一是键值对形式的复杂属性，由单个条件决定属性值，称为“Key-Value”(KV)结构。如“申请”类的“操作说明”，不同入口对应不同的操作说明；二是由多个条件组合才能决定属性值的属性，称为“Compound Value Type”(CVT)结构，如需要考虑退货原因、商品、有效期等多种因素才能确定用户购买的商品能不能退。

4.3.4 本体模型表示

汇总前 3 个步骤的成果，形成搭建基本框架阶段的完整本体模型记录表，包括类表和属性表、类表记录类及类的层次关系、属性表记录属性的定义及属性所属的类。

4.4 融合知识文档和用户数据完善知识结构

搭建基本框架阶段获得由 FAQ 中抽象出的类及类的属性，形成基本框架。基本框架需要融合其他层次知识进行完善。知识文档代表业务运营方提供的领域知识，用户问题等代表从用户角度需求的领域知识。本阶段融合知识文档、用户问题两类数据完善知识结构。

4.4.1 融合知识文档为完善知识结构

知识文档包含更全面的领域知识，FAQ 与知识文档存在交叉重合的关系。以知识文档为补充完善知识结构旨在从知识文档中抽象出 FAQ 没有包含的类及属性，完善知识结构。经过基本框架的搭建，本体模型构建者积累了对领域知识的认知。此时，从长篇知识文档中识别出基本框架未包含的类及属性具备可行性。与搭建基本框架步骤相同，采用手工或自动方法，从知识文档中识别出基本框架尚未包含的类及属性，添加到基本框架中。

选择退换货流程、退换货政策等知识文档，识别知识文档中含有而 FAQ 中没有涉及的类或属性，添加到基本框架中。

4.4.2 融合用户数据为完善知识结构

用户真实问题等直接反映用户需求，支持本体模型与用户真实问题的范围与粒度对齐，确保覆盖率，细化粒度。从智能问答或其他渠道收集的用户问题往往体量较大，可以首先采用机器学习等方法对用户问题进行相似性聚类，并标记每一类中的代表性数据。人工审阅每类中的代表性数据，从中识别出新的类和属性，添加到基本框架中。融合用户问题数据完善知识结构采取与“搭建基本框架”同样的步骤来完成，首先获取领域术语，再定义类和属性。

在退换货领域，选取一段时间内的相关数据，采用 Bert 模型对数据进行分类，并标记每类中最具有代表性的数据。从代表性数据中识别基本框架中未包含的类或属性。如在用户问题中有关于如何取消申请或修改申请的问题，需要在“申请”类下添加“取消申请”“修改申请”子类，如表 1 所示：

表 1 类和类的层次结构－以用户问题完善

顶类	一级类	二级类	三级类	……
Thing				
	Process			
		申请		
			取消申请	
			修改申请	
		审核		
……	……	……	……	

4.5 融合业务系统数据对齐知识结构

业务系统中的动态数据记录当下用户最新、最详细的信息，智能问答与业务系统的实时交互将有利于智能应答实时根据最新信息，提供细粒度的答案或解决方案。因此，有必要融合业务系统数据进一步完善本体模型。

在对齐数据结构过程中，首先识别数据结构中的术语是否需要新增为类或属性。其次根据类和属性的定义方法，定义类或属性，与本体模型中原有类及属性进行融合、映射。本体模型与业务系统数据结构的融合与映射主要存在 4 种情况：①一对一映射，业务系统数据结构与本体都存在含义相同的字段或属性，此时仅需将业务系统字段与本体属性映射一对一映射，如订单编号；②多对一映射，业务系统的一个字段如果包含了本体多个属性，需要将业务系统中的字段的值拆分为多段，分别与本体中的属性对应，如业务系统中的“退货状态”字段对应本体中各个流程如“申请”“审核”类的“状态”；③一对多映射，业务系统中的多个字段，对应本体中的一个属性，此时需要将多个字段对应到本体的一个属性，例如本体模型中的“日期”对应业务系统中的“年、月、日”3 个字段；④业务系统数据结构中的字段本体中尚未涵盖，此时需要在本体的相应位置添加类或属性。在与业务系统数据结构对齐过程中需注意始终以应用目标为导向。

与退换货业务相关的业务系统有企业退换货系统、电商商城等业务系统。将知识图谱本体模型与业务系统中的退换货流程、订单等数据结构对齐，在本体模型中添加“订单”、“退货申请单”等类，支持在用户提问时访问业务系统获取实时数据，根据用户当下的具体情况给出答案或引导操作。如用户在询问“退货到什么阶段了？”时，可实时查询并返回业务系统中用户退货业务当前的状态。

在所有循环及细节调整完成后，知识图谱的本体模型形成稳定版本，支持后续基于本体模型的实例层构建。

4.6　退换货知识图谱本体模型

经过搭建基本框架、完善知识结构、对齐知识结构等基于多层次数据的循环构建流程后，获得退换货本体模型如图 2 所示：

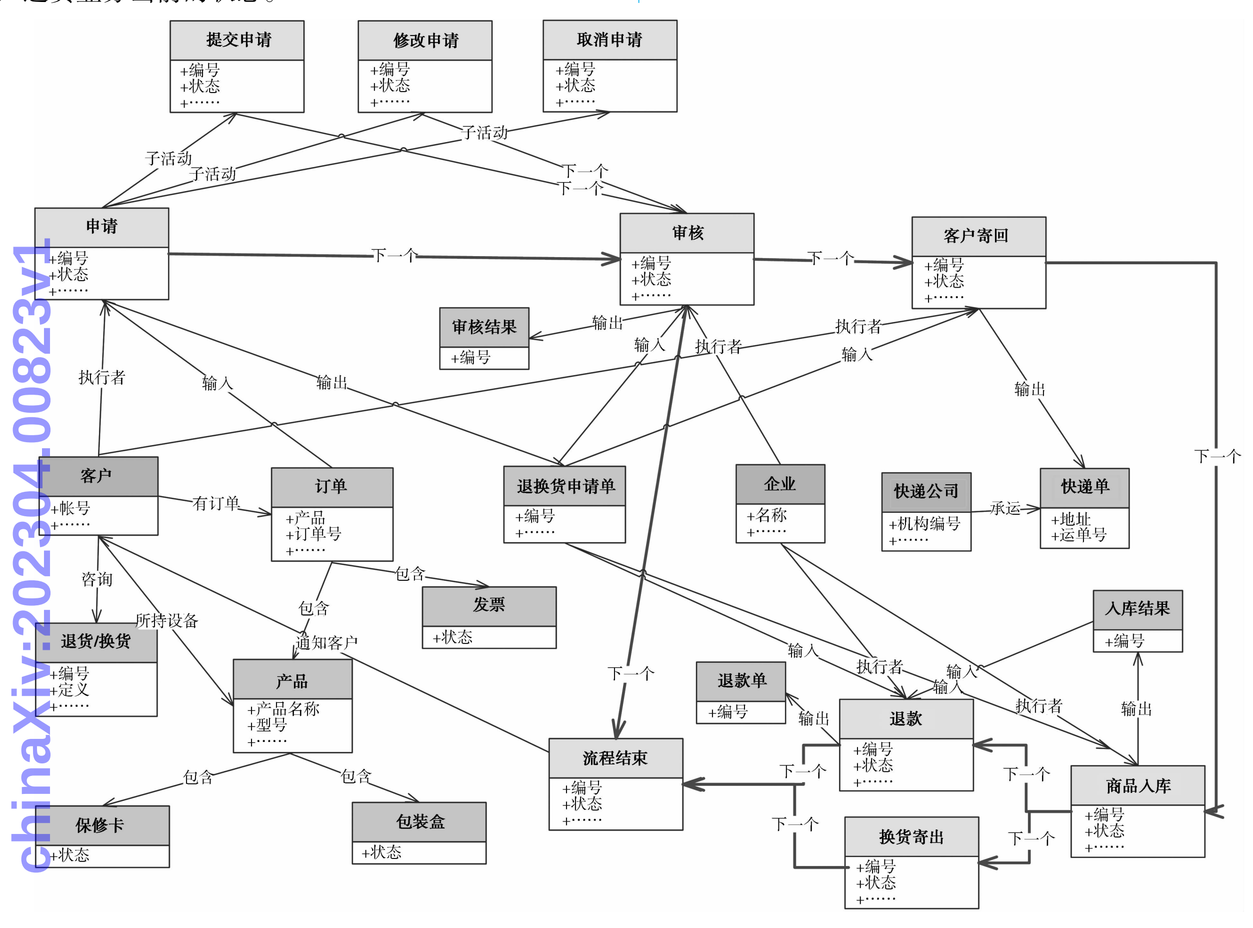


图 2　退换货本体模型

5　退换货本体模型应用效果评估

以退换货本体模型为模式层，构建退换货知识图谱实例层，形成退换货知识图谱。为评估基于退换货知识图谱的智能问答解决现有困境的效果，本研究对相关数据进行统计分析或试验测评，主要统计指标有知识的覆盖率、细粒度以及智能问答的准确度、解决率。

知识覆盖率方面，在知识图谱构建之前退换货领域仅有 50 余个问答对，知识图谱构建完成后其属性就已接近 250 个，属性值或属性值组合可覆盖上千个问题，覆盖率大幅提升。

细粒度方面，知识图谱能够灵活支撑回答不同细粒度的用户问题。一方面，由于融合了用户问题、业务系统数据等多层次数据，知识图谱本体模型类及属性的设置与用户问题的粒度对齐。基于知识图谱的智能问答不再是概括性问答对，而是能针对用户的具体情况调用或生成答案；另一方面，知识图谱本体模型的多级结构，使得智能问答不仅能回答最细粒度的用户问题，也可以应对用户粗粒度的问题，如退换货本体模型中“退货”类下有“线上退货”“线下退货”两个子类，当用户询问“实体店买的商品如何退货”时，对应子类“线下退货”，当用户询问“如何退货”时，对应父类“退货”。

智能问答的准确度和解决率方面，将退换货知识图谱部署到智能问答系统中，随机抽取人工在线客服的400余个原始退换货相关问题，采用人工评测方式检测退换货知识图谱部署前后的智能问答准确率和解决率。测试结果显示，退换货知识图谱应用后准确率提升50%，解决率提升300%，效果显著。具体如表2所示：

表2　基于知识图谱的智能问答试验结果

场景	准确率/%	解决率/%
退换货知识图谱应用前	58.5	21
退换货知识图谱应用后	88	84

统计分析与试验结果显示，在原有企业提供知识的基础上，融合用户数据、业务系统动态数据等多层次数据构建知识图谱本体模型，对于提高基于知识图谱的智能问答的能力效果显著。一方面是因为融合用户数据使本体模型的结构粒度与真实应用环境中的用户需求对齐，知识覆盖率与应用需求无限接近；另一方面，与业务系统数据的对齐后，本体模型支持通过获取实时动态数据定位用户的具体情况，为用户提供更精准的问题答案，而非概述性的答案线索，进一步细化可回答的问题粒度。融合多层次数据使得基于知识图谱的智能问答准确率、解决率大幅提升。

6　结论

本文针对当前智能问答面临的问题，提出融合多层次数据来构建知识图谱本体模型的方案，以支持基于问答对的智能问答转向基于知识图谱的智能问答。在案例分析中，本文将构建的退换货知识图谱本体模型应用于智能问答系统中进行试验。结果显示，融合多层次领域数据的知识图谱本体模型有效确保了智能问答中知识的覆盖率和细粒度，大幅提高了智能问答准确率和解决率，实现了智能问答能力升级。

此外，融合多层次数据的本体模型作为领域全景知识结构，不仅可支持知识图谱模型构建，还可应用于更多场景。例如，融合用户数据后可发现企业提供知识的不完整，反推业务数据的优化，也可支持构建领域知识分类体系和意图分类体系。

本研究主要考虑企业业务领域本体的构建，融合的多层次数据是基于企业业务领域的情况分析而获取。不同的领域的多层次数据类型可能存在差异，此方法在其他类型领域的应用还需在今后的工作中进一步研究验证。

参考文献：

[1] SINGHAL A. Official Google blog: introducing the knowledge graph: things, not strings[EB/OL]. [2021－07－02]. https://www.blog.google/products/search/introducing-knowledge-graph-things-not/.

[2] IRENE P. Knowledge graphs vs property graphs a brief overview and comparison[EB/OL].[2021－05－21]. https://www.topquadrant.com/wp-content/uploads/2020/08/Knowledge-Graphs-vs.-Property-Graphs-A-Brief-Overview-and-Comparison.pdf.

[3] GRUBER T R. A translation approach to portable ontology specifications[J]. Knowledge acquisition, 1993, 5(2): 199－220.

[4] GRÜNINGER M, FOX M S. Methodology for the design and evaluation of ontologies[EB/OL]. [2021－12－21]. http://citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.44.8723&rep=rep1&type=pdf.

[5] USCHOLD M, KING M. Towards a methodology for building ontologies[M]. Edinburgh: University of Edinburgh, 1995.

[6] FERNÁNDEZ-LÓPEZ M, GÓMEZ-PÉREZ A, JURISTO N. Methontology: from ontological art towards ontological engineering[J]. Engineering workshop on ontological engineering,1997(Mar.):33－40.

[7] SURE Y, STAAB S, STUDER R. On-to-knowledge methodology (OTKM)[M]//Handbook on ontologies. Berlin:Springer, 2004: 117－132.

[8] SUÁREZ-FIGUEROA M C, GÓMEZ-PÉREZ A, FERNÁNDEZ-LÓPEZ M. The NeOn methodology for ontology engineering [M]//Ontology engineering in a networked world. Berlin:Springer, 2012: 9－34.

[9] DE NICOLA A, MISSIKOFF M, NAVIGLI R, et al. A software engineering approach to ontology building[J]. Information systems, 2009, 34(2): 258－275.

[10] NOY N F, MCGUINNESS D L. Ontology development 101: a guide to creating your first ontology[EB/OL]. [2021－12－21]. https://corais.org/sites/default/files/ontology_development_101_aguide_to_creating_your_first_ontology.pdf.

[11] GANGEMI A, PRESUTTI V. Ontology design patterns[M]//Handbook on ontologies. Berlin:Springer, 2009: 221－243.

[12] PRESUTTI V, DAGA E, GANGEMI A, et al. eXtreme design with content ontology design patterns[C]//Proceedings of workshop on ontology patterns. Washington D. C.:CEUR-WS. Org., 2009: 83－97.

[13] 朱妍昕,徐维,王霞,等.基于FAIR原则的循证医学文献本体构建——以哮喘药物治疗文献本体为例[J/OL].情报理论与实践:1－12[2021－09－30]. http://kns.cnki.net/kcms/detail/11.1762.G3.20210819.1201.002.html.

[14] 贾君枝.《汉语主题词表》转换为本体的思考[J].中国图书馆学报,2007(4):41－44.

[15] 段瑞龙,宋文.国内外叙词表转换本体方法研究综述[J].情报杂志,2012,31(7):66－71.

[16] ZHEN Z, TANG A M, SHEN J Y, et al. Thesaurus-based approach for building domain ontology with a case study of military aircraft prototype ontology construction[J]. Journal of Southeast University (English Edition), 2006(3): 353 – 356.
[17] WIŚNIEWSKI D, POTONIEC J, ŁAWRYNOWICZ A, et al. Analysis of ontology competency questions and their formalizations in SPARQL-OWL[J]. Journal of Web semantics, 2020(29): 1 – 13.
[18] 杨波,廖怡茗. 面向企业动态风险的知识图谱构建与应用研究[J]. 现代情报,2021,41(3):110 – 120.
[19] 张肃,许慧. 基于知识图谱的企业知识服务模型构建研究[J]. 情报科学,2020,38(8):68 – 73.
[20] DE NICOLA A, MISSIKOFF M. A lightweight methodology for rapid ontology engineering[J]. Communications of the ACM, 2016, 59(3): 79 – 86.
[21] SUÁREZ-FIGUEROA M C, GÓMEZ-PÉREZ A. Ontology requirements specification[M]//Ontology engineering in a networked world. Berlin: Springer, 2012: 93 – 106.
[22] D'ANTONIO F, MISSIKOFF M, TAGLINO F. Formalizing the OPAL eBusiness ontology design patterns with OWL[M]//Enterprise interoperability II. London: Springer, 2007: 345 – 356.

作者贡献说明:
周毅:论文框架设计,模型构建,论文撰写及修改;
刘峥:提供论文选题和框架设计,模型构建,论文修改;
粟小青:参与模型构建和评审,论文修改;
金体成:负责数据收集,参与模型构建,进行应用试验。

Ontology Model Construction of Question-Answering Knowledge Graph Integrating Multi-Level Data

Zhou Yi[1] Liu Zheng[1,2] Su Xiaoqing[3] Jin Ticheng[3]
[1] National Science Library, Chinese Academy of Sciences, Beijing 100190
[2] Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190
[3] Huawei Device Co. Ltd., Shenzhen 518129

Abstract: [**Purpose/significance**] Aiming at problems of intelligent Q&A based on Q&A pairs such as low accuracy and resolution rate and poor user satisfaction, this paper constructs a knowledge graph (KG) ontology model that supports the realization of dynamic and accurate intelligent Q&A based on the knowledge graph. [**Method/process**] First, the paper analyzed the current problems and causes of intelligent question answering, and proposed a plan to build a knowledge graph to support intelligent question answering. Second, On the basis of existing ontology model construction methods, the paper proposed a multi-round loop method integrating multi-level data, which used the business data provided by the enterprises, user data and business system dynamic data as the data sources. And the core steps were to build a basic framework, improve the knowledge structure, and align three cycles of the knowledge structure. Finally, this paper took the domain of return and exchange as a case to describe the concrete steps of ontology model construction, from zero, added incrementally, and constructed ontology model of knowledge graph. [**Result/conclusion**] This paper applies the knowledge graph with the return ontology model as the schema layer in an intelligent Q&A system for testing. The evaluation results show that the accuracy rate increased by 50% and the precision rate increased by 300% after the return and exchange knowledge graph is online. So, the proposed ontology model construction method sorts out the complete and fine-grained domain knowledge structure from scattered domain knowledge, can provide accurate answers to users in intelligent Q&A, and can effectively solve the intelligent Q&A dilemma based on Q&A pairs.

Keywords: knowledge graph ontology model accurate Q&A multi-level data